Paired Exome Analysis of Barrett's Esophagus and Adenocarcinoma

Supplementary Information

Paired Exome Analysis of Barrett's Esophagus and Adenocarcinoma

Index

Paired Exome Analysis of Barrett's Esophagus and Adenocarcinoma

Supplementary Table 1: Patient, Barrett's and Tumor Characteristics in the 25 paired samples

Supplementary Table 2: List of tumor suppressors and their given coverage

Supplementary Table 3: Patient and tumor demographics for laser capture microdissection samples

Supplementary Table 4: Tumor suppressor multiplicity

Supplementary Table 5: Tumor suppressor gene and pathway loss

Supplementary Table 6: EAC gene alterations

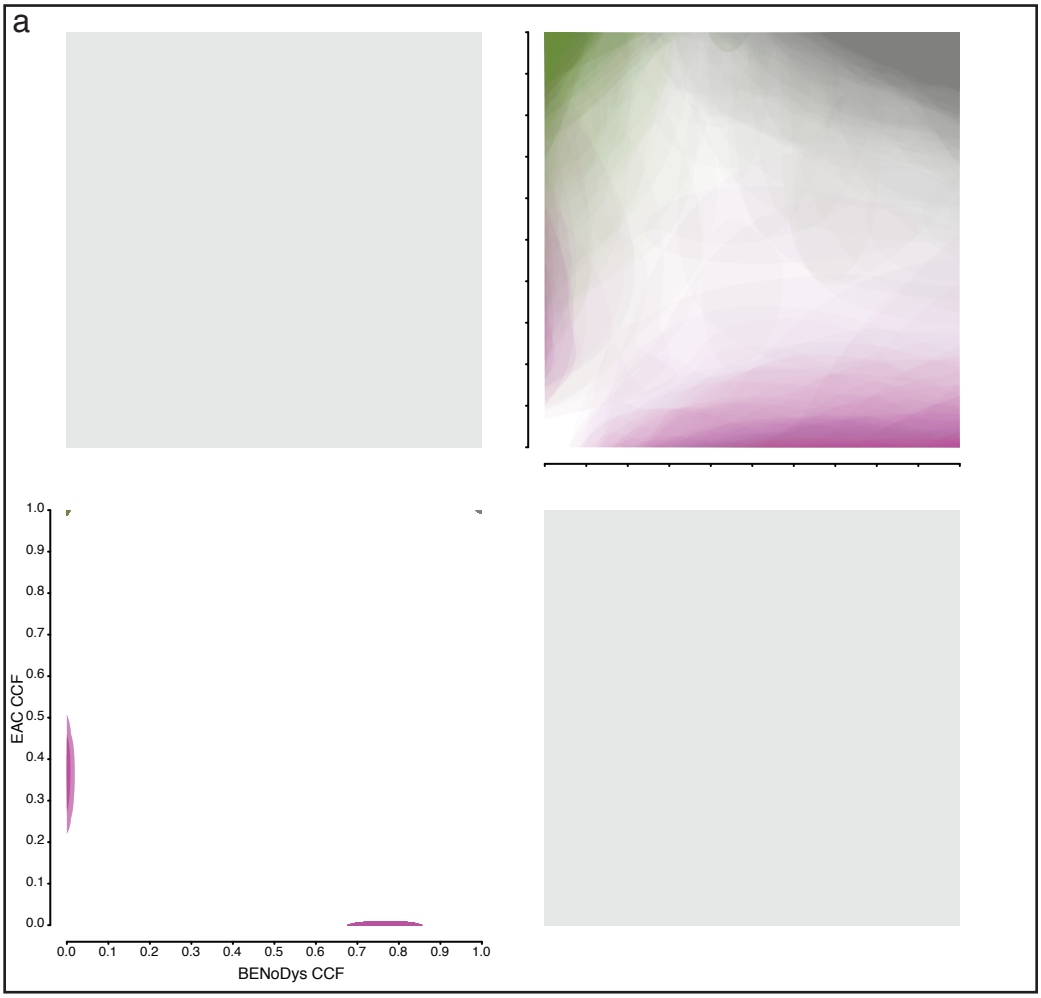
Paired Exome Analysis of Barrett's Esophagus and Adenocarcinoma

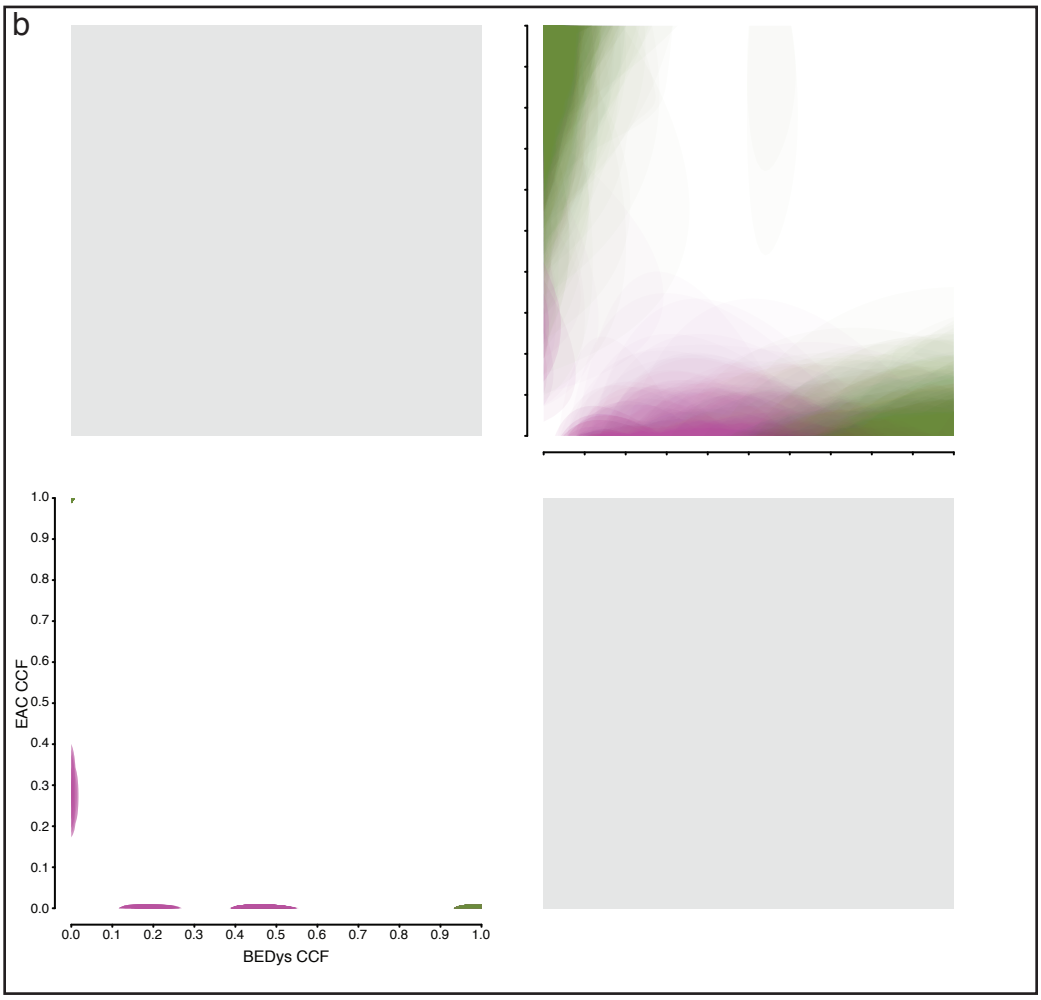
Supplementary File 1: Frozen Trio (normal, BE, EAC) MAFs

Supplementary File 2: Frozen Trio ABSOLUTE plots

Supplementary File 3: FFPE multi-sample MAFs

Supplementary File 4: FFPE multi-sample ABSOLUTE plots

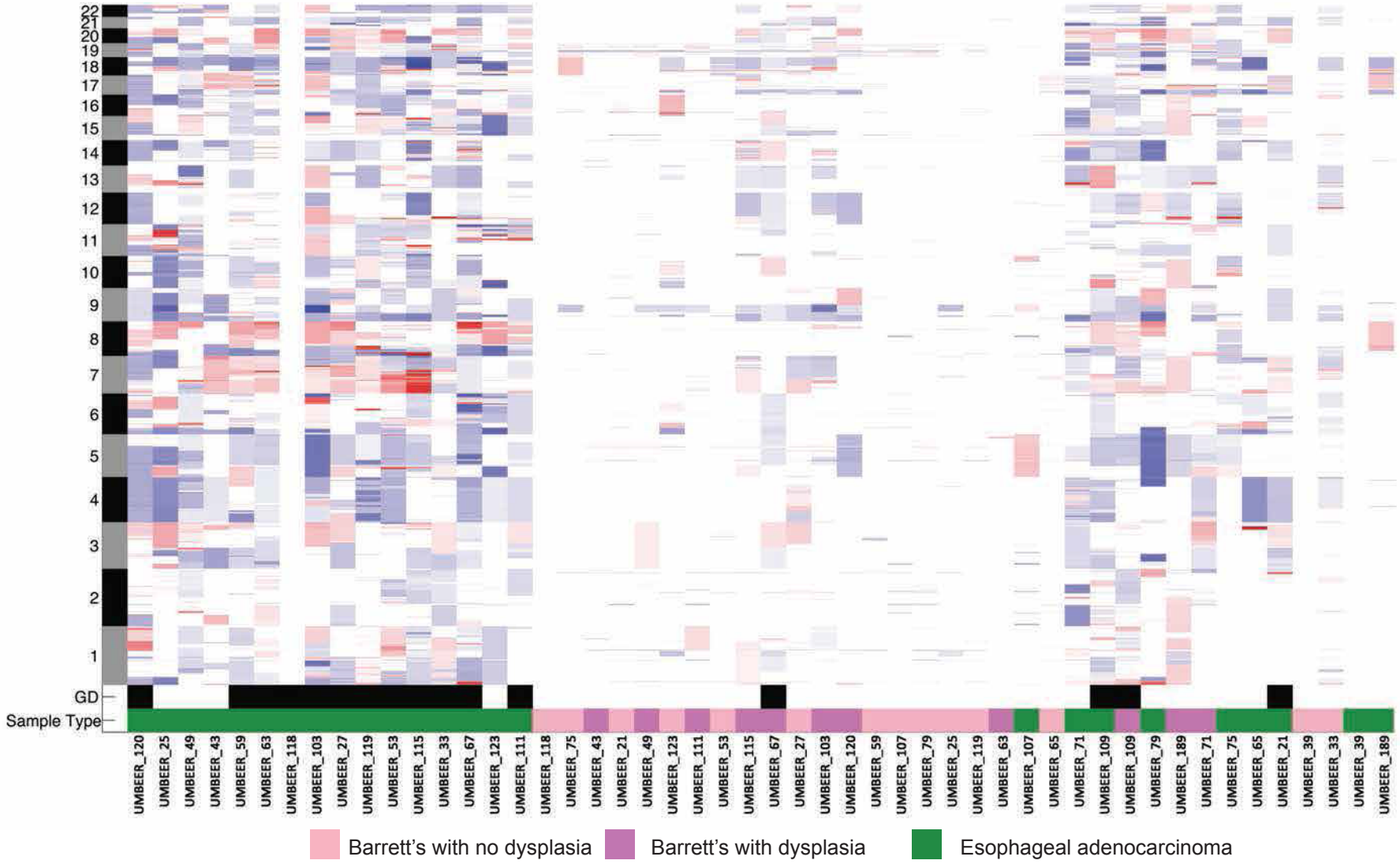




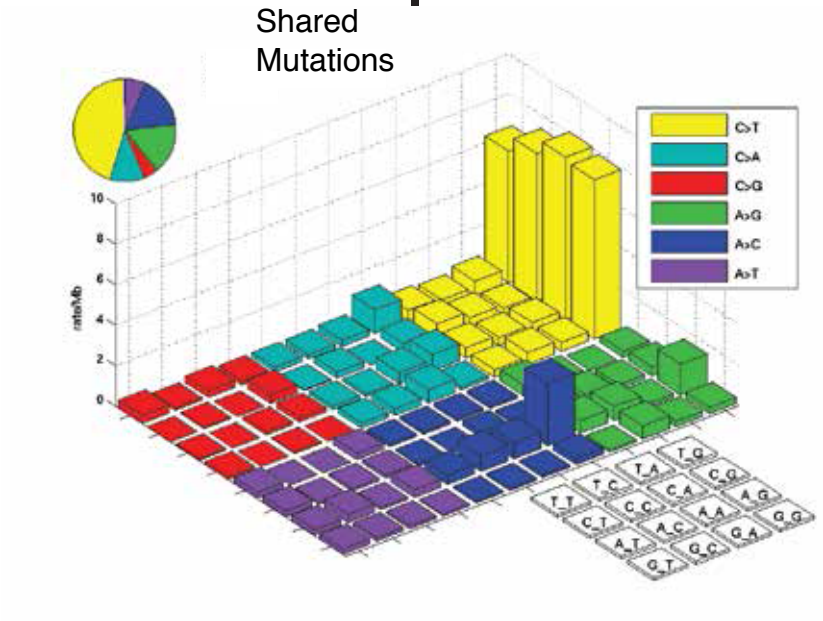
Supplementary Figure 1: 2D mutation-CCF plots for clonally related and unrelated BE/EAC pairs.

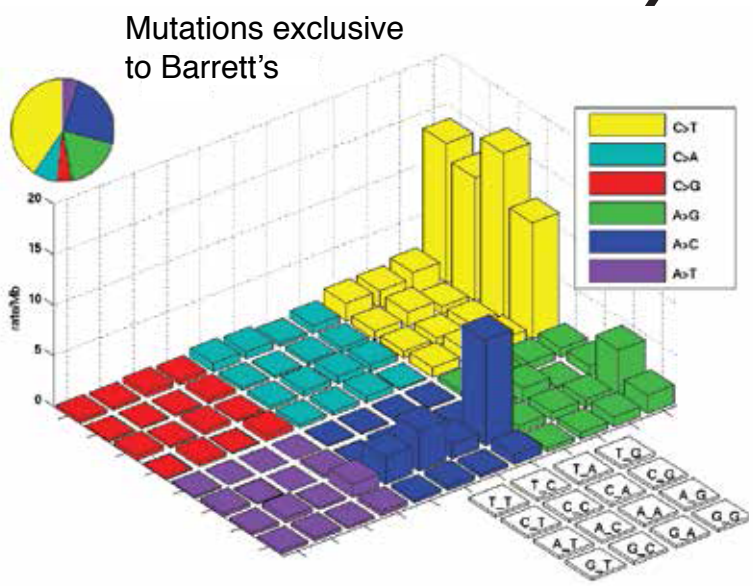
Supplementary Figure 1: 2D mutation-CCF plots for clonally related and unrelated BE/EAC pairs.

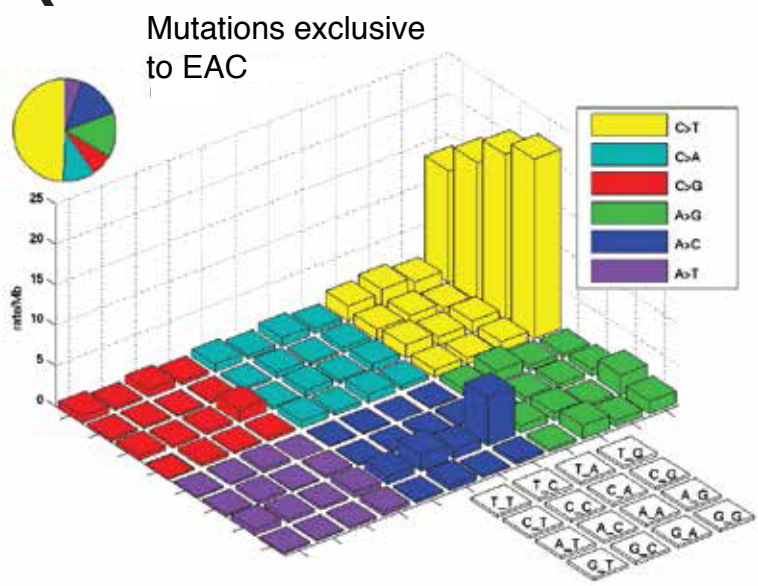
a, b Cancer-cell fraction (CCF) probability distributions for BE samples (*x*-axis) vs. paired EAC samples (*y*-axis) are shown for mutations detected in each pair. The lower-left plots show the result of the 2D Bayesian clustering algorithm applied to the input CCF distributions in the upper-right. CCF-distributions for each mutation are displayed as transparent filled-contours corresponding to density equal to scales of 0.99, 0.95, 0.8, 0.5, and 0.1 times each density at the mode. The degree of transparency is inversely proportional to each density at the mode (i.e., mutations with more uncertain CCF have greater transparency). Mutations are colored according to the expected value of their post-clustering CCF density, as follows: grey – CCF ≥0.8 in both samples; green – CCF ≥ 0.8 in one sample and uncalled in the other; pink – CCF < 0.8 in one sample and uncalled in the other; orange: CCF < 0.8 one sample and called in the other. For clarity, only mutations covered by ≥ 20 sequencing reads are displayed. We note that, for the purposes of this plot, small numbers of reads supporting the alternate allele for each mutation were discarded if they were insufficient to result in a mutation call using MuTect (i.e., mutations were not force-called).
a Case #33 (clonally related)
b Case #43 (clonally unrelated)
Similar plots for all 30 patients are available in **Supplementary File 3**.



Supplementary Figure 2: Barrett's and esophageal adenocarcinoma samples from the same patient do not co-cluster
Hierarchical clustering of paired EAC and BE samples using CNVs. EACs and BE predominately cluster with each other with only one pair (sample 109) clustering together.

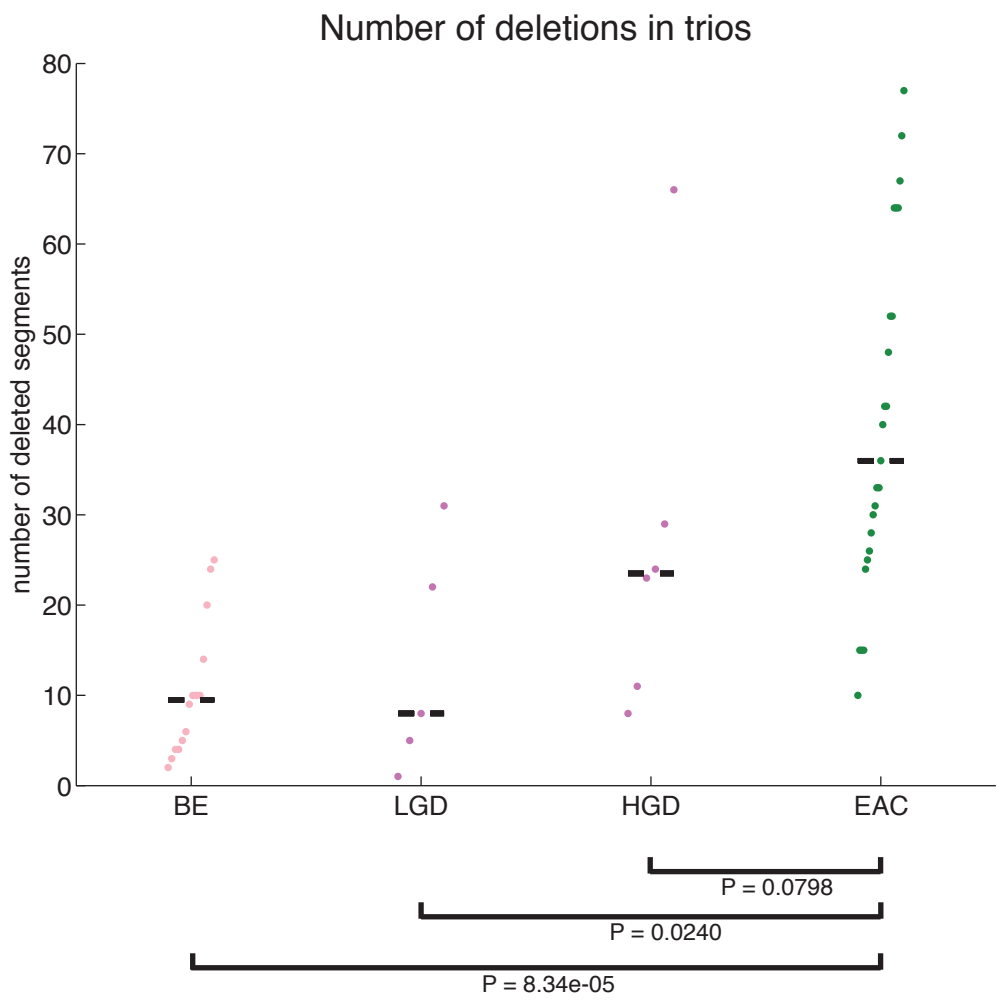




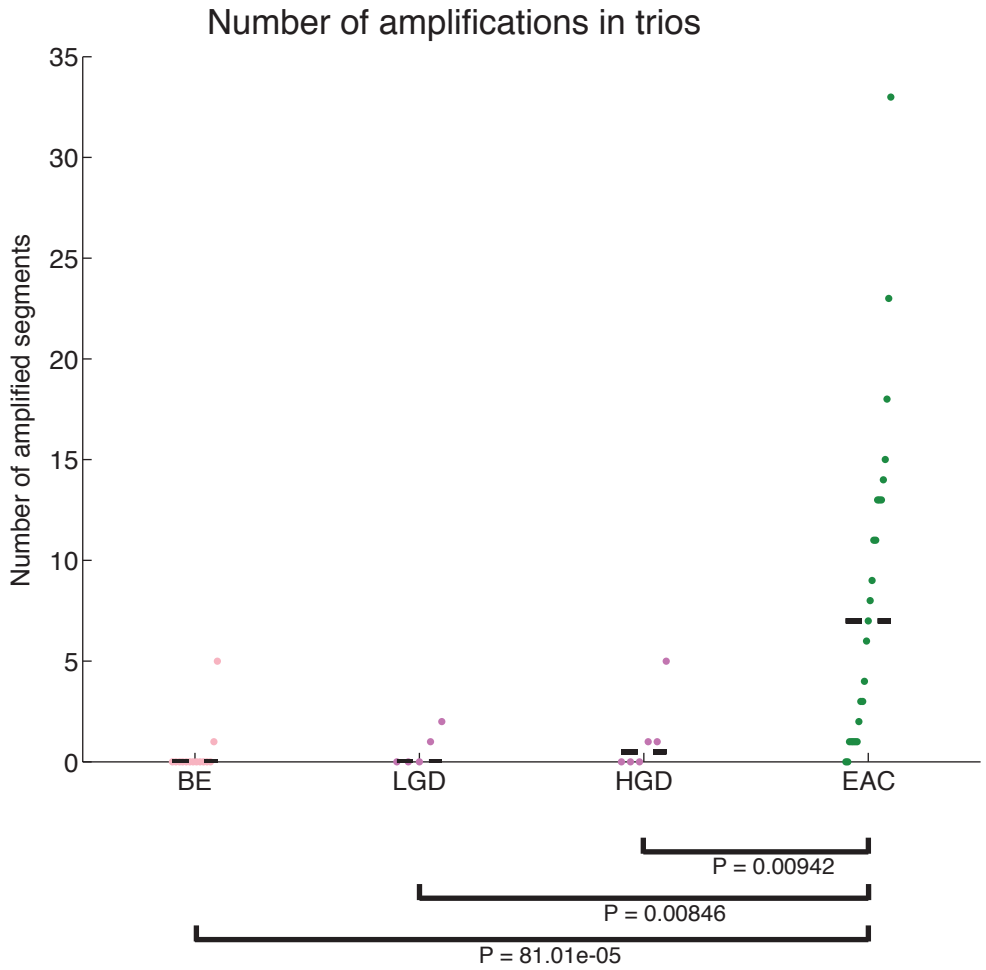


Supplementary Figure 3: Barrett's and esophageal adenocarcinoma share an A to C mutation signature
"Lego" plots of mutation frequencies across 25 paired Barrett's and esophageal adenocarcinoma samples. Mutations shared between the paired samples are shown on the top, while mutations exclusive to the Barrett's sample in the bottom left and mutations exclusive to the esophageal adenocarcinoma sample shown on the bottom right. Base substitutions are divided into six categories to represent the six possible base changes (each category represented by a different color). Substitutions are further divided by the 16 possible flanking nucleotides surrounding the mutated base as listed in the corresponding box legend. The inset pie chart indicates the distribution of all mutations for a given middle mutated base.

a



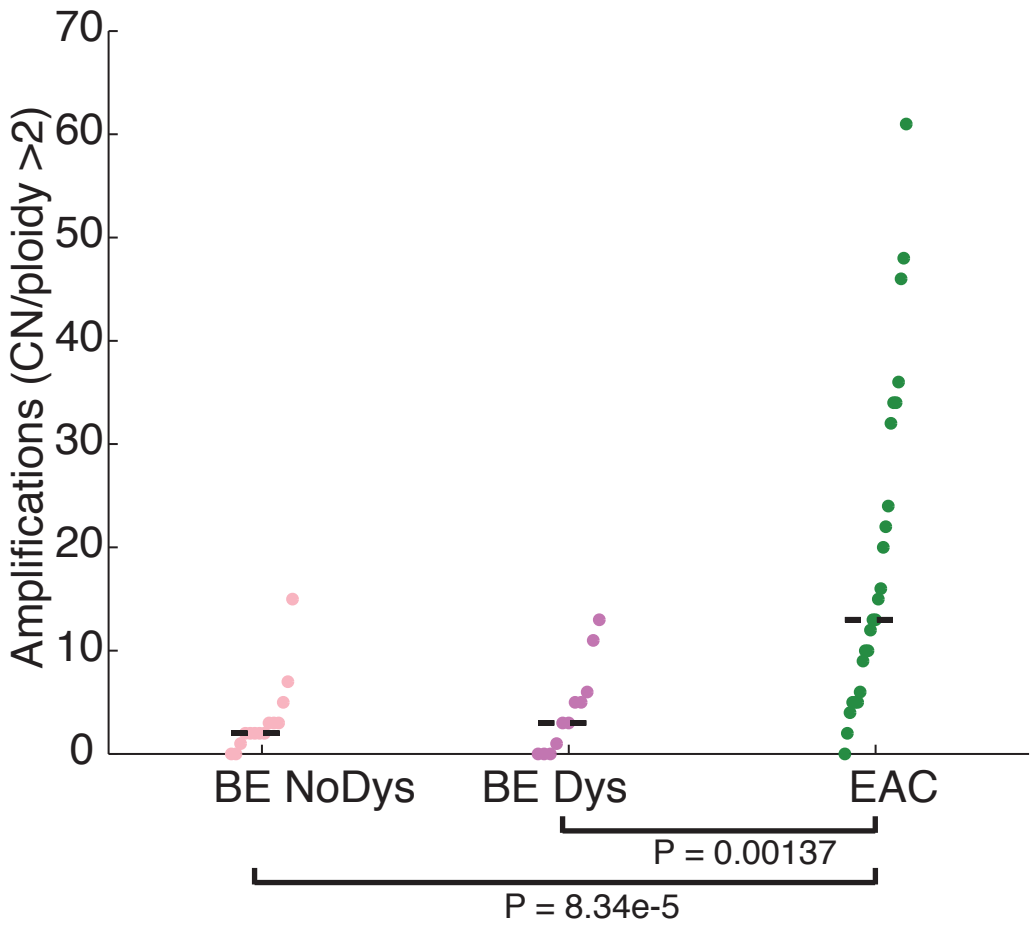
b



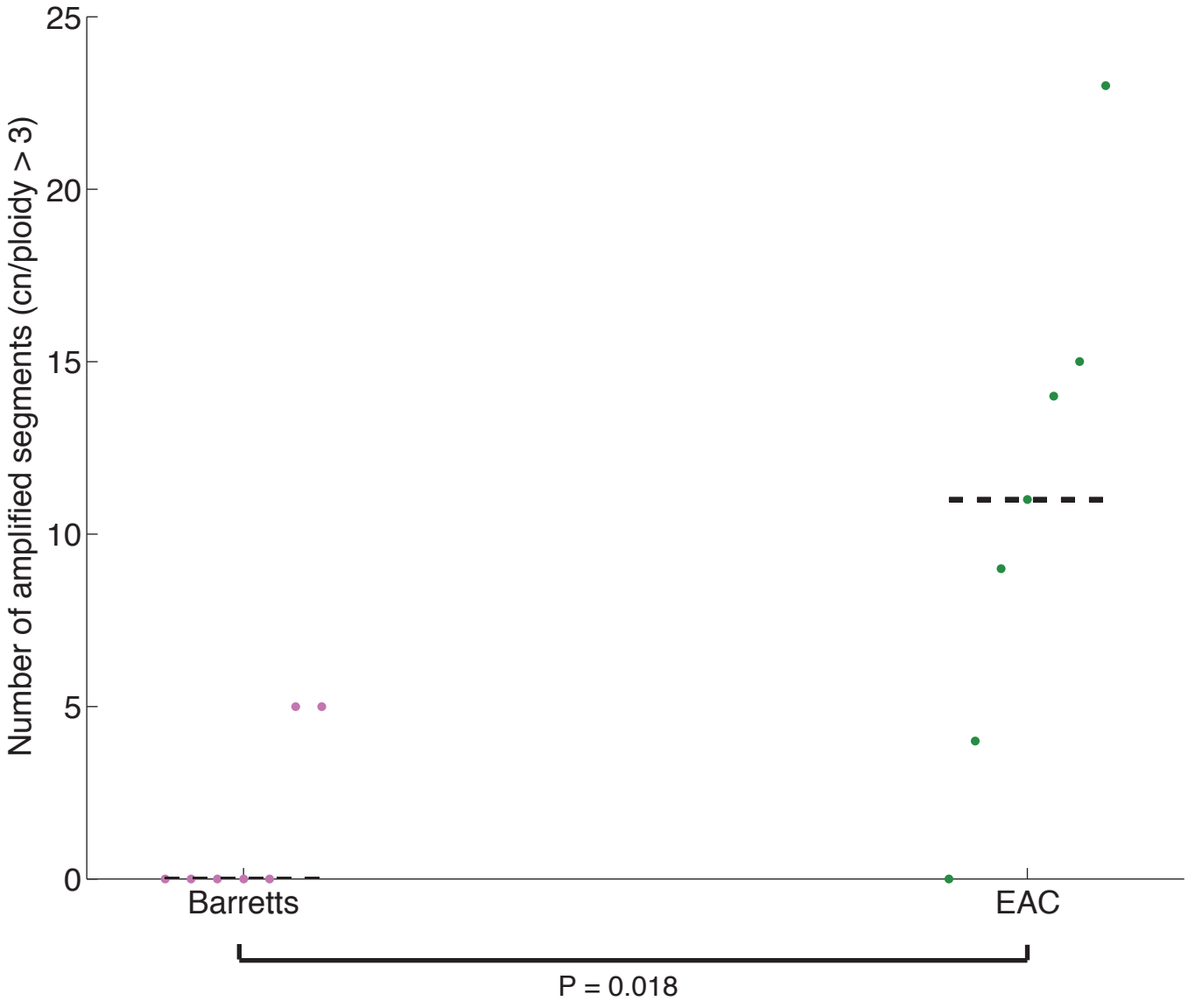
Supplimentary Figure 4: Number of deletions and amplifications within Barrett's, Barrett's with low grade dysplasia, Barrett's with high grade dysplasia, and esophageal adenocarcinoma.

Supplementary Figure 4: Number of deletions and amplifications within Barrett's, Barrett's with low grade dysplasia, Barrett's with high grade dysplasia, and esophageal adenocarcinoma

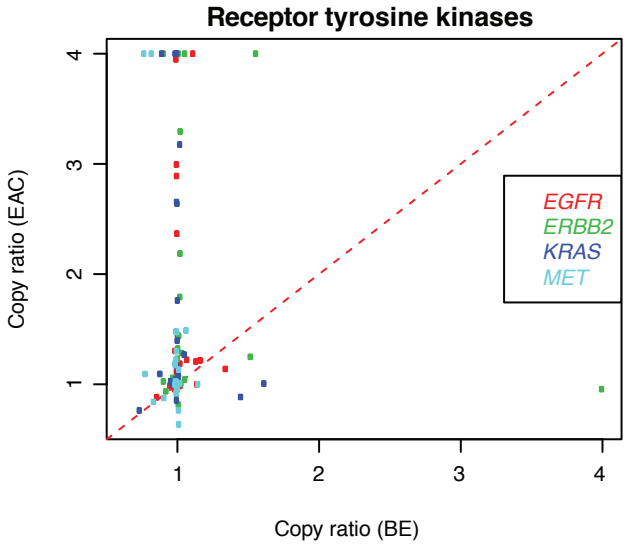
Copy number plots showing the number of A) deletions and B) amplifications within each sample. BE samples are categorized based upon the pathologic determination from hematoxylin and eosin stained frozen sections as either Barrett's with no dysplasia (BE), Barrett's with low grade dysplasia (LGD), or Barrett's with high grade dysplasia (HGD). Line represents the mean. CN: copy number.

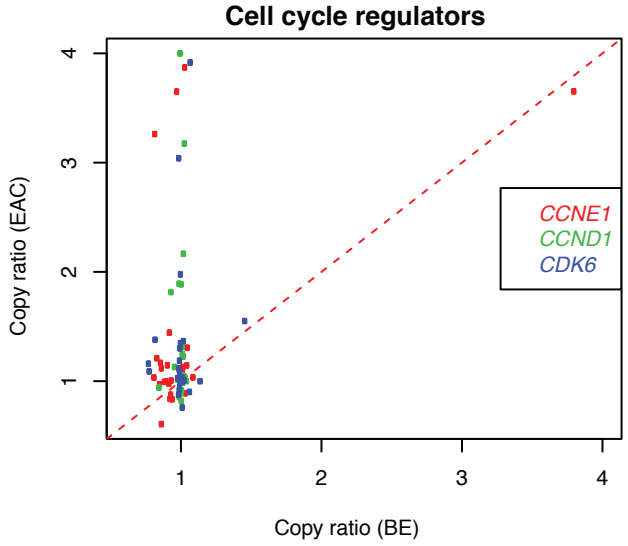


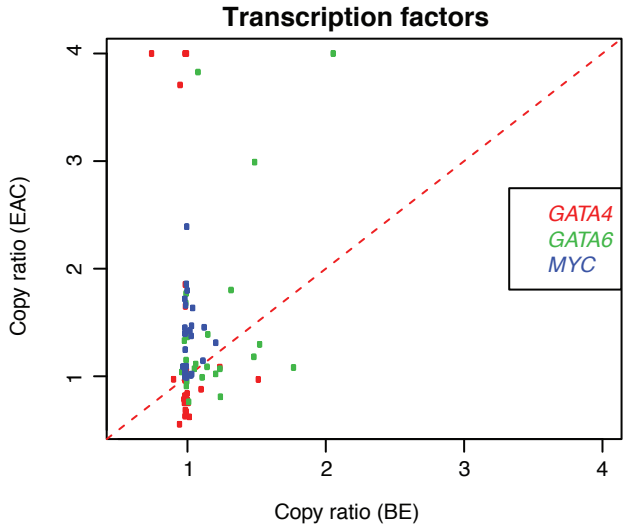
Supplementary Figure 5: Number of amplifications in Barrett's and esophageal adenocarcinoma using lower cutoff value.
Number of amplified segments per sample for Barrett's with no dysplasia (BENoDys), Barrett's with dysplasia (BEDys), and esophageal adenocarcinoma using the less stringent copy number cutoff of copy number (CN) /ploidy > 2 (4 copies for a 2N cell). Line represents the mean.



Supplementary Figure 6: Number of amplifications in the seven most highly related Barrett's and esophageal adenocarcinomas.
Number of amplified segments per sample for Barrett's and esophageal adenocarcinoma (EAC) using only the 7 most highly related pairs. Line represents the mean.



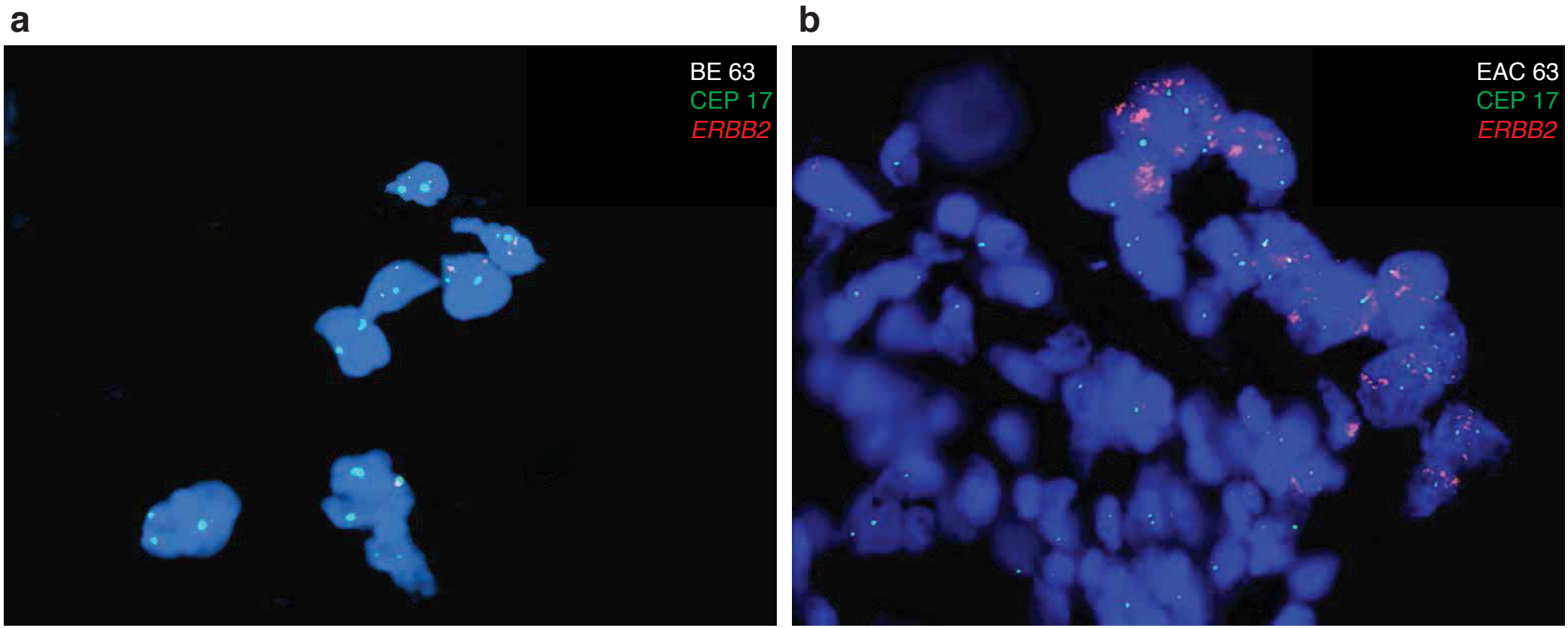




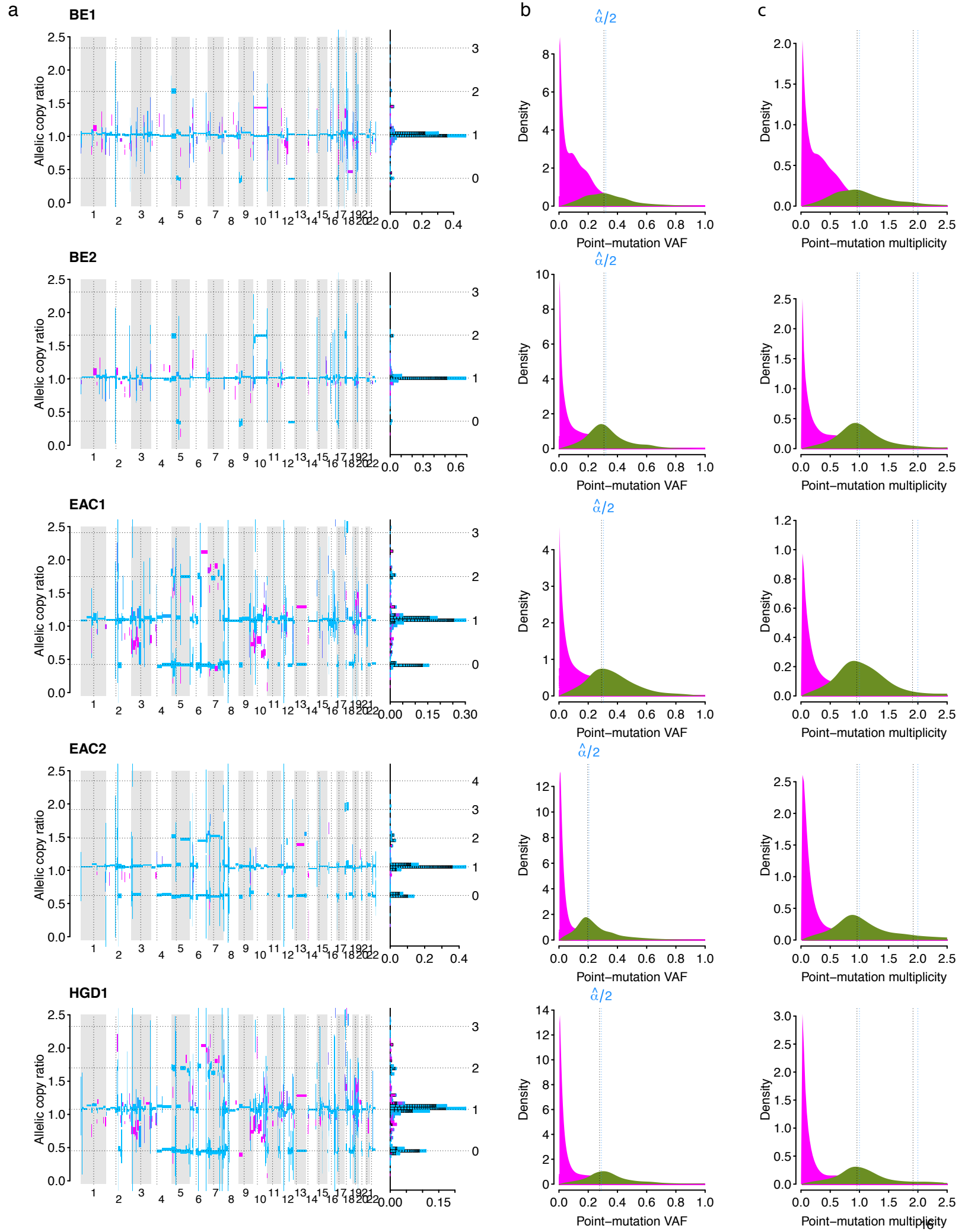
Supplementary Figure 7: Barrett's esophagus lacks amplification of oncogenes commonly identified in esophageal adenocarcinoma

Supplementary Figure 7: Barrett's esophagus lacks amplification of oncogenes commonly identified in esophageal adenocarcinoma

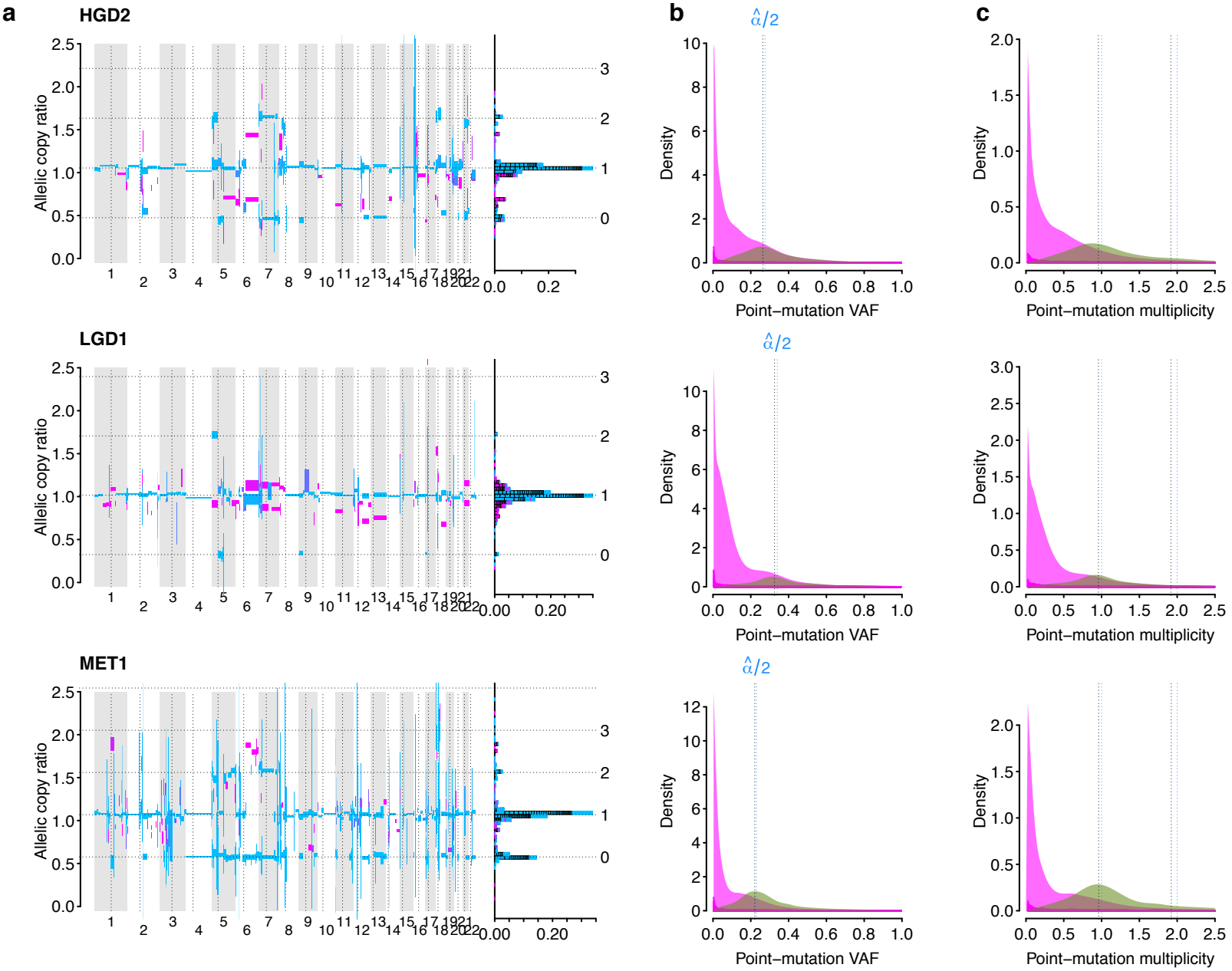
Dot plot of oncogene copy ratio of paired BE and EAC samples. Gene copy ratio of commonly amplified oncogenes are shown with paired BE on the x-axis and EAC on the y-axis. Graph capped at a copy ratio of 4.



Supplementary Figure 8: Fluorescent in-situ hybridization for ERBB2 in patient 63.
Genomic amplification of ERBB2 was identified in sample EAC 63 but not in BE 63. These results were confirmed by FISH analysis with an average ERBB2/CEP17 ratio of 1.07 (negative) in BE (A.) and 3.33 (positive) in EAC (B.).



Supplementary Figure 9: Results of ABSOLUTE on samples from patient P7



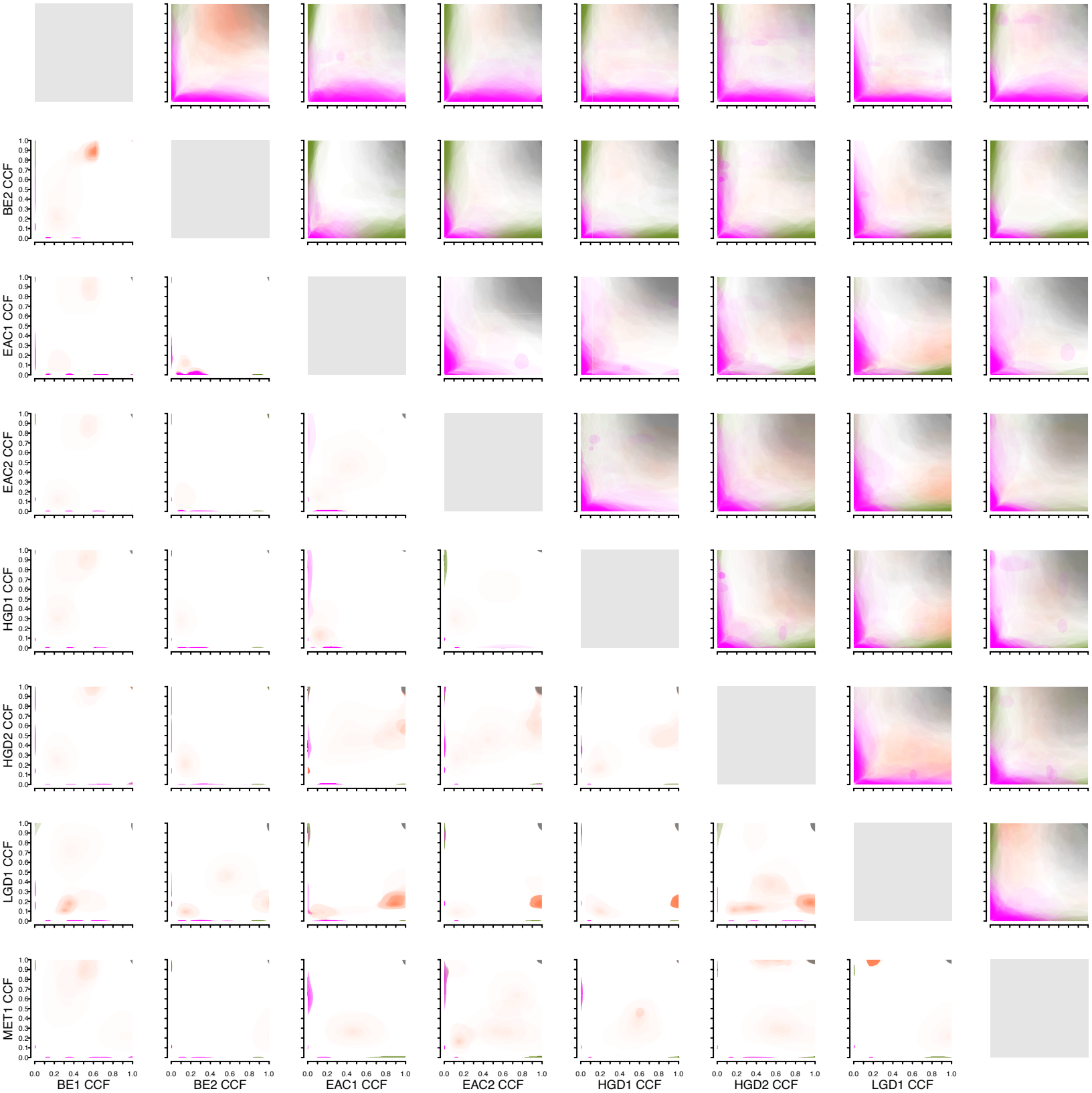
Supplementary Figure 9: Results of ABSOLUTE on samples from patient P7

a. Allelic segmental copy-ratios (y-axis) for each pair of homologous chromosomes across the entire genome (x-axis). The height of each segment indicates the 95% confidence interval for the copy-ratio. Dotted horizontal lines correspond to modeled absolute allelic copy-numbers, as indicated on the right-hand y-axis of the adjacent summary histogram (right). Segments are colored as blue if the copy-ratios are consistent with all neoplastic cells contributing to the sample having the same copy-number in that region (CCF=1), or pink if they appear to derive from a heterogeneous mixture of copy-numbers (CCF<1).

b. The summation of probability distributions (y-axis) over the fraction of alternate reads (VAF; x-axis) for the somatic mutations detected in the sample is shown as filled transparent curves for mutations with CCF=1.0 (green) and with CCF < 1.0 (pink). The dotted vertical line indicates the inferred sample purity divided by 2. For samples with fewer than 500 mutations detected, the distributions for individual mutations are plotted using solid lines.

c. Distributions over mutation VAF are rescaled using ABSOLUTE to correct for estimated sample purity and local copy-number (displayed as in b). This yields distributions over multiplicity (x-axis), the expected number of mutant alleles present per cell in the sampled cancer-cell population. Mutations with multiplicity < 1 are restricted to a subset of this population (CCF < 1.0). Mutations with multiplicity = 2 likely occurred by whole genome doubling, or by chromosomal gain including the mutant allele.

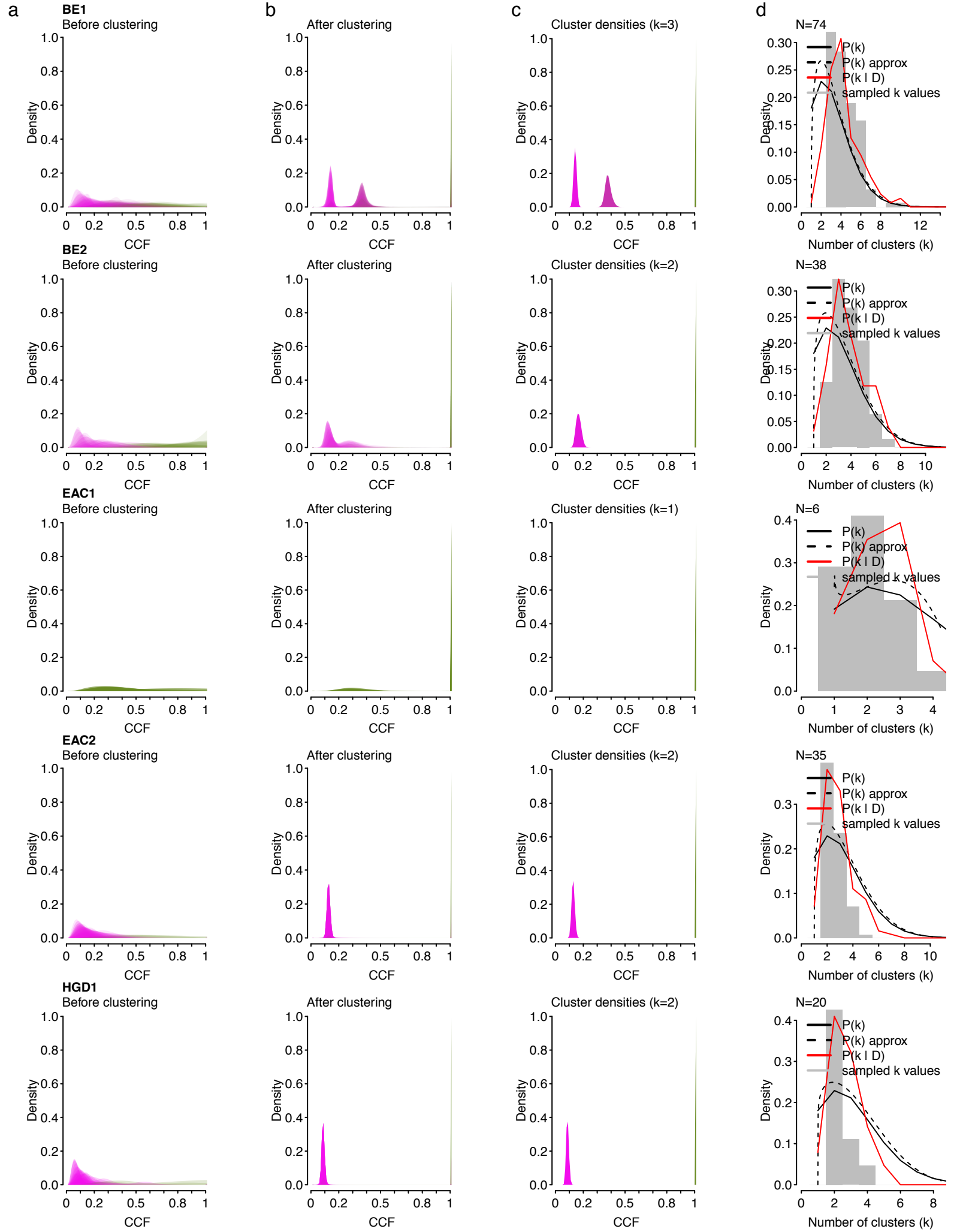
Similar plots for all 30 patients are available in Supplementary File 2, 4.



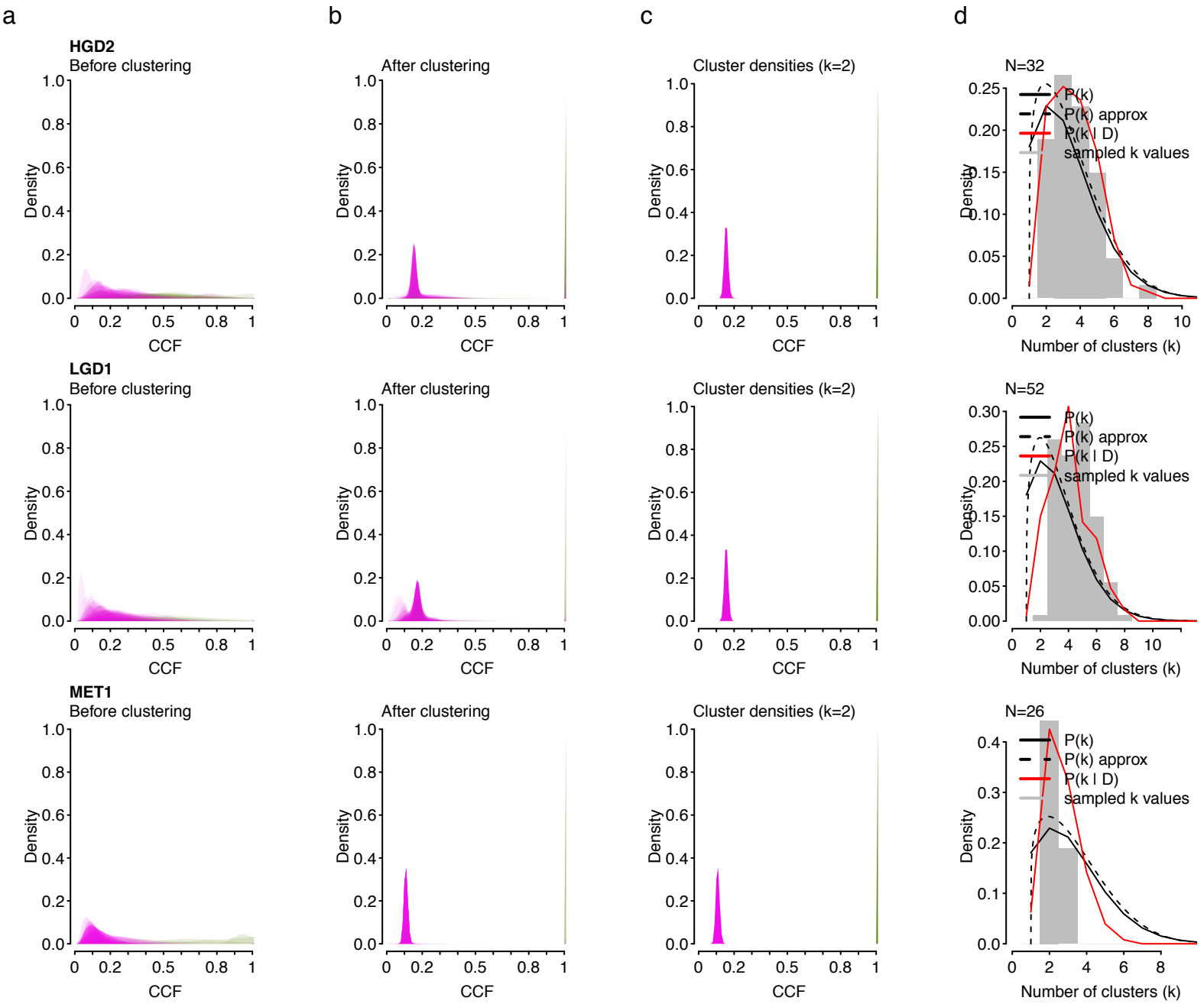
Supplementary Figure 10: 2D Bayesian clustering analysis of point-mutation CCF distributions in patient P7

Two-dimensional cancer-cell fraction (CCF) probability distributions for each unique pair of samples from patient 7 are shown for mutations detected in each pair. The plots on the lower-left (lower diagonal matrix) show the result of the 2D Bayesian clustering algorithm applied to the input CCF distributions, shown in the upper-triangular matrix (each plot in the lower-triangle corresponds to its transposed coordinates in the upper-triangle; note that tissue-sample labels refer only to plots in the lower-triangle.) CCF-distributions for each mutation are displayed as in **Supplementary Figure 1**.

Similar plots for all 30 patients are available in **Supplementary File 2, 4**.



Supplementary Figure 11: Bayesian clustering of private point-mutations CCF distibutions in patient P7



Supplementary Figure 11: Bayesian clustering of private point-mutation CCF distributions in all sequenced tissue-samples from patient P7

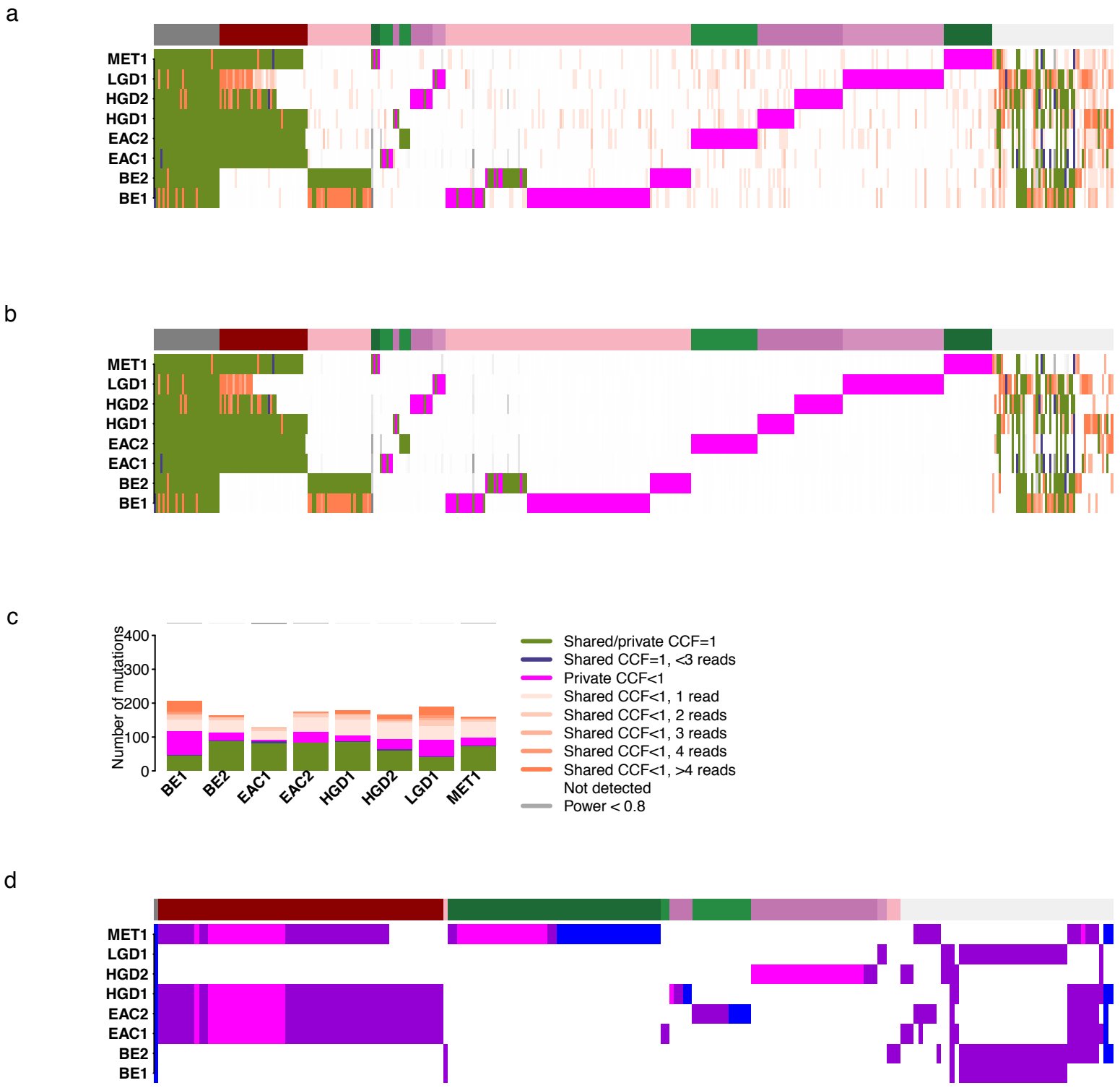
a. Pre-clustering probability distributions over CCF (from ABSOLUTE) are plotted for each mutation as transparent filled curves with color determined by hard-cluster assignment (shown in C).

b. Post-clustering probability distributions over CCF are shown for each mutation.

c. Posterior densities of the hard-clusters.

d. prior and posterior distributions over k (cluster number) are shown, as well as the histogram of sampled k values from the MCMC run (after discarding the burn-in samples).

Similar plots for all 30 patients are available in **Supplementary File 2, 4**.



Supplementary Figure 12: Genetic alterations supporting phylogeny construction in patient P7

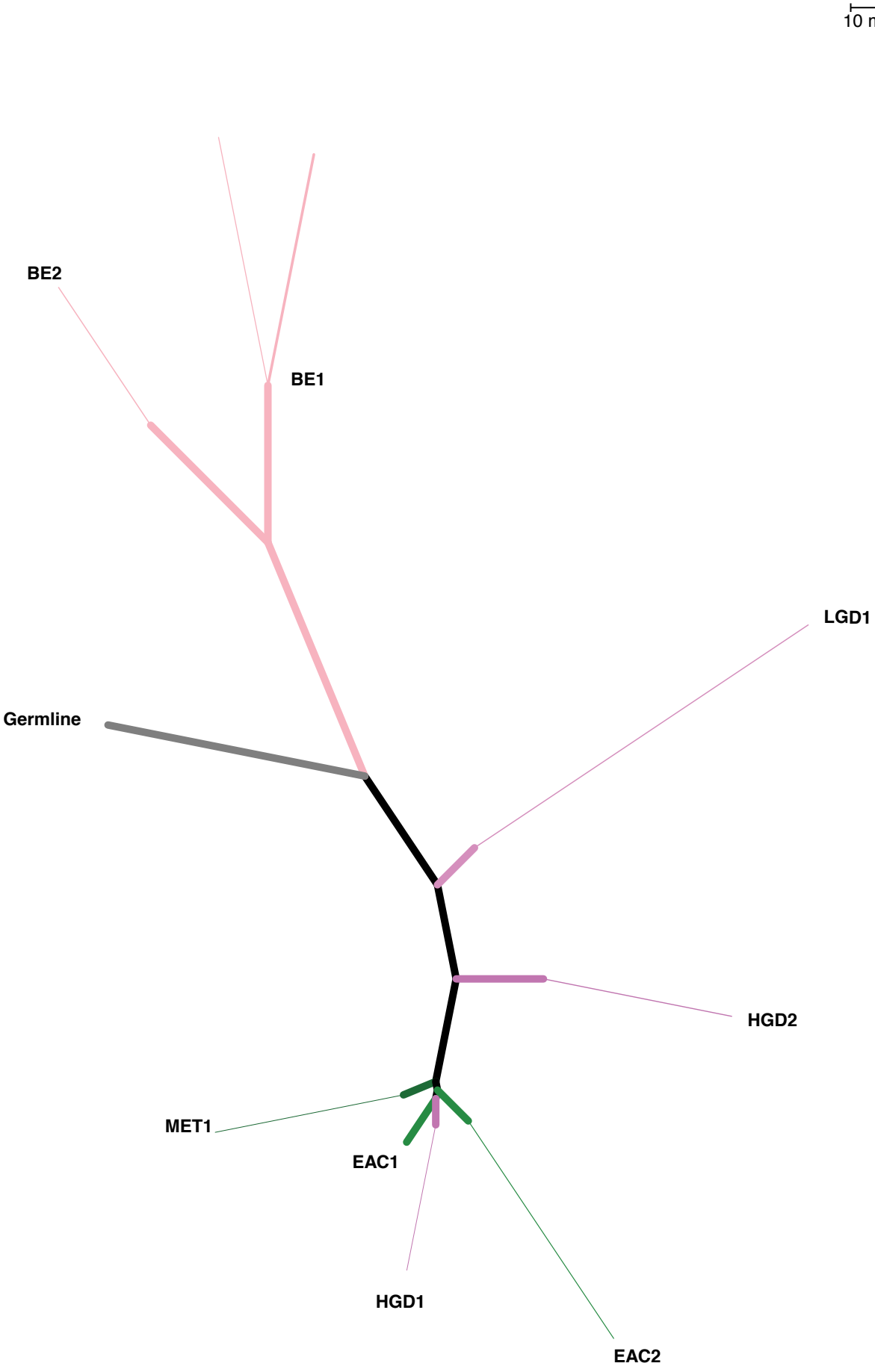
a. The matrix of all point mutations (SSNVs and indels; columns) detected in any of the tissue samples from this case (rows) is shown. The mutations are colored according to the legend in **c**, with the exception of undetected mutations, which are colored according to a linear color-scale indicating paired-detection power ranging from black (power=0) to white (power > 0.99). The color bar on the top of the matrix indicates the assignment of mutations to branches of the phylogenetic tree (shown in **Supplemental Fig. 13**). Mutations on the far right (light grey) were not assigned to a branch of the tree.

b. The data from **a** is shown after removing mutations rejected by the automatic phylogeny inference procedure (i.e., mutations present in > 1 sample with any CCF value < 1.0 and fewer than 3 supporting reads).

c. Stacked bar-plot of the data in **a**, summarized over each of the tissue samples from this case.

d. The matrix of all SCNAs, called at the gene level, (columns) detected in any of the tissue samples from this case (rows) is shown. The color bar at the top indicates phylogenetic branch assignment, as in **a**.

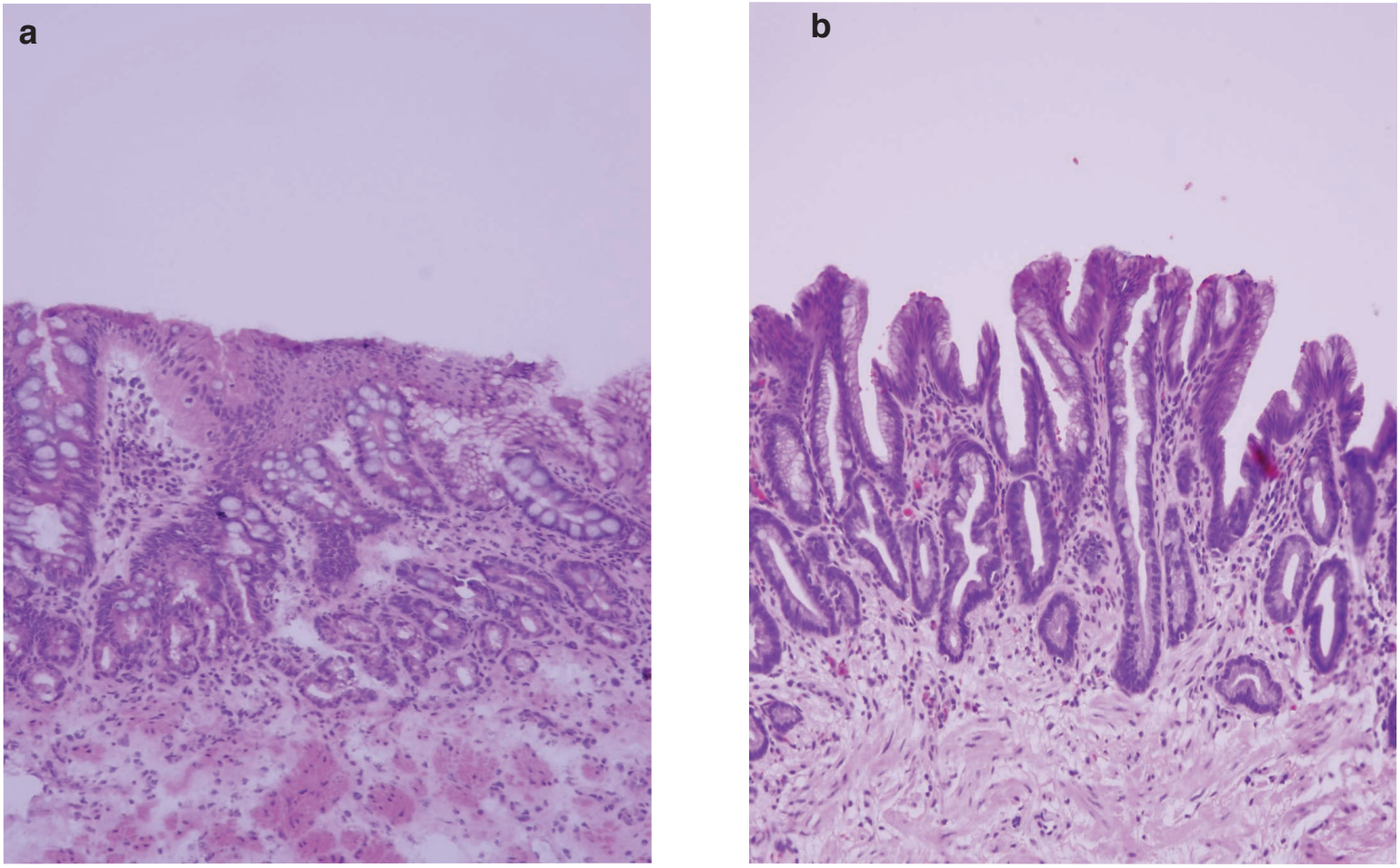
Similar plots for all 30 patients are available in **Supplementary File 2, 4**.



Supplementary Figure 13: Phylogenetic tree for patient P7

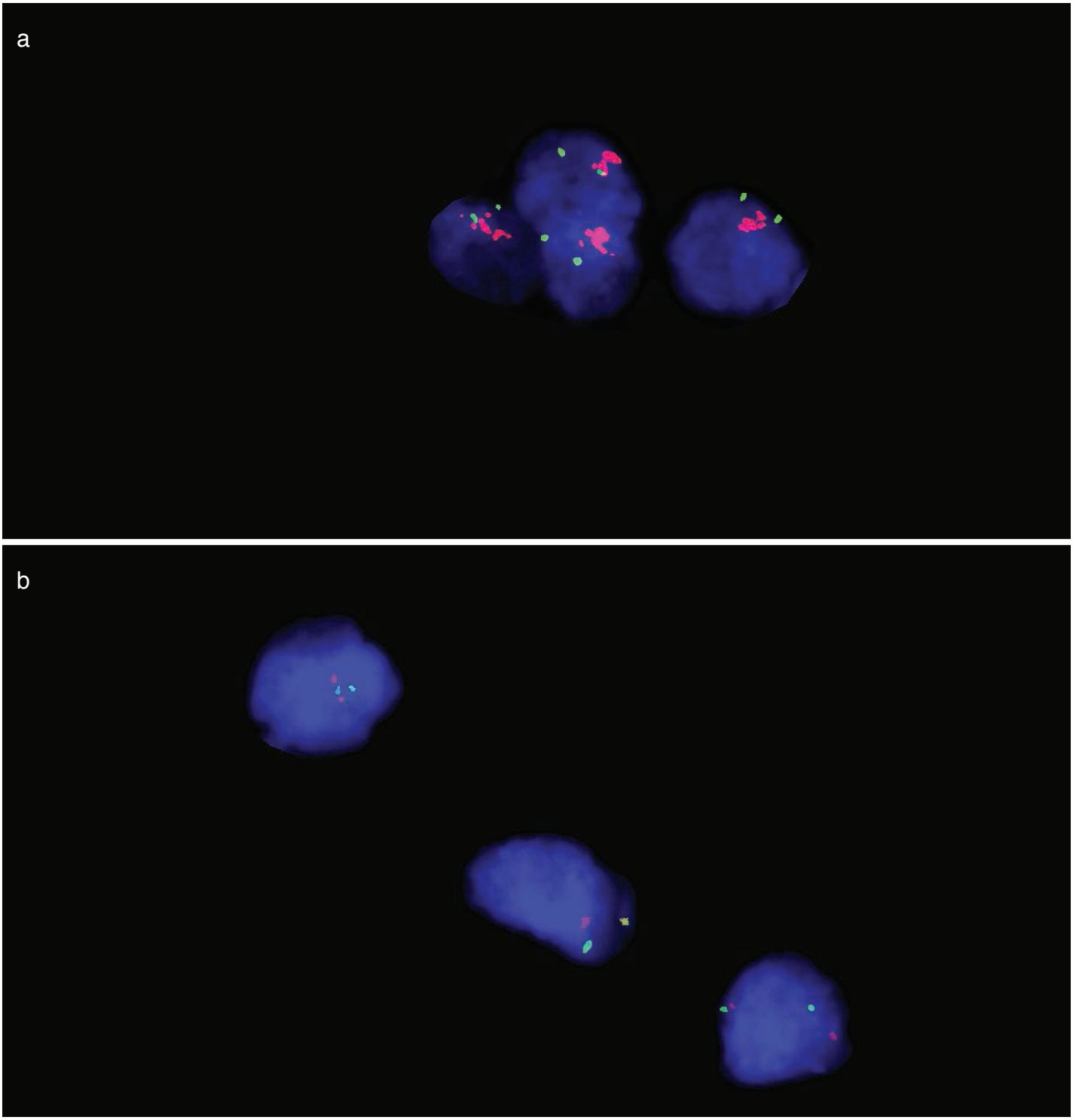
Supplementary Figure 13: Phylogenetic tree for patient P7
The automatically generated phylogenetic tree obtained by maximum parsimony on the matrix of binary mutation presence / absence data (**Supplemental Fig. 12a**,**b**) is shown. This tree was generated assuming the branched-sibling model. The manually adjusted tree correcting for partially shared minor subclones (CCF < 1) between some samples is shown in **Figure 3**. The thickness of each line corresponds to the CCF of the mutations on that branch. For each tissue sample, the longest sub-branch is labeled with the name of the tissue sample.

Similar plots for all 30 patients are available in **Supplementary File 2, 4**.

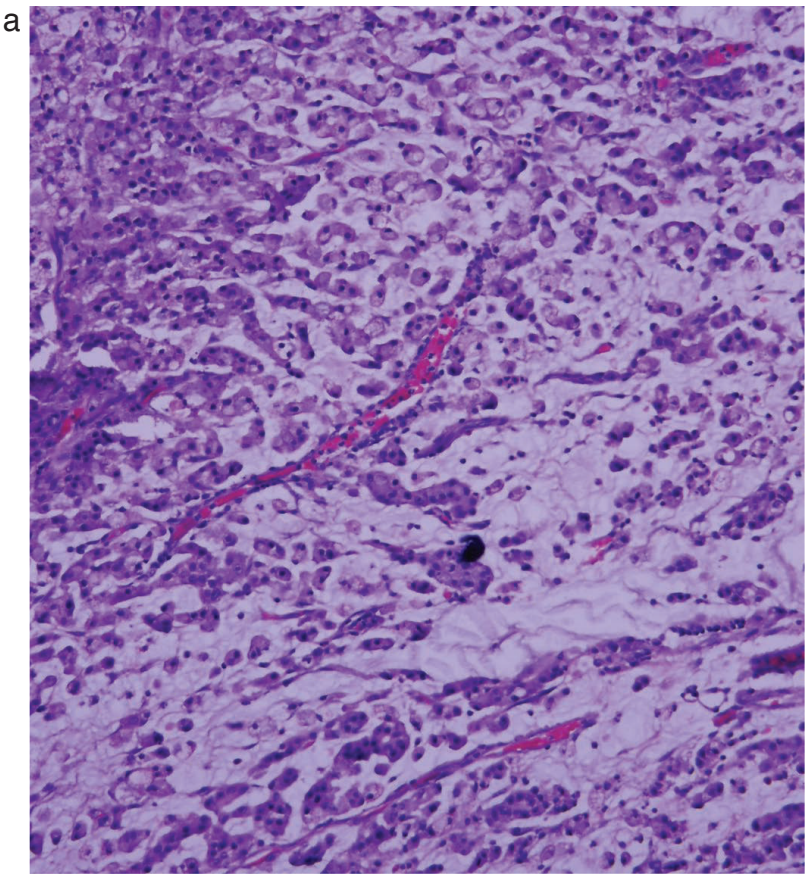


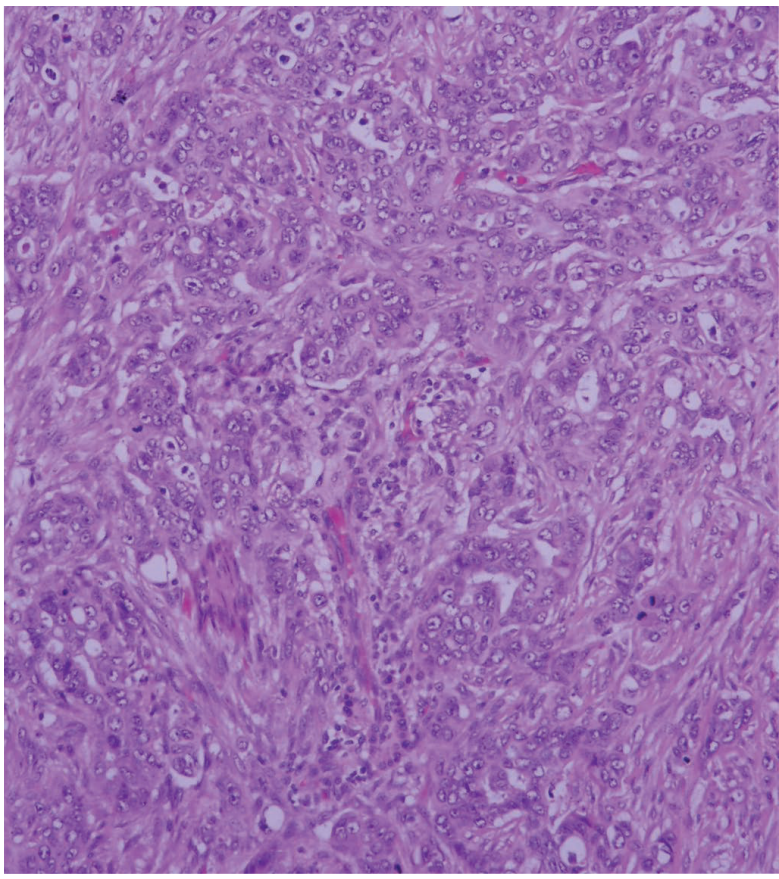
Supplementary Figure 14: Hematoxylin and Eosin stained sections of non-dysplastic Barrett's which harbor TP53 mutations

Representative hematoxylin and eosin stained a) frozen and b) formalin fixed paraffin embedded sections showing Barrett's esophagus. Laser capture microdissection with whole exome sequencing revealed both contained TP53 mutations.



Supplementary Figure 15: Representative images of *KRAS* fluorescent in-situ hybridization. *KRAS* FISH of patient 7 sample EAC1 (**a**.) showing high level amplification and sample HGD2 (**b**.) showing no amplification. *KRAS* probe is red while CEP 12 is green.





EAC1
- *CCND1* Amplification
- Non-genome doubled
- Calculated ploidy: 1.93

Average FISH centromeric probe count

CEP 2	CEP 4
1.74	1.86

EAC2
- *MET, MYC*, and 3q26 Amplification
- Genome doubled
- Calculated ploidy: 3.07

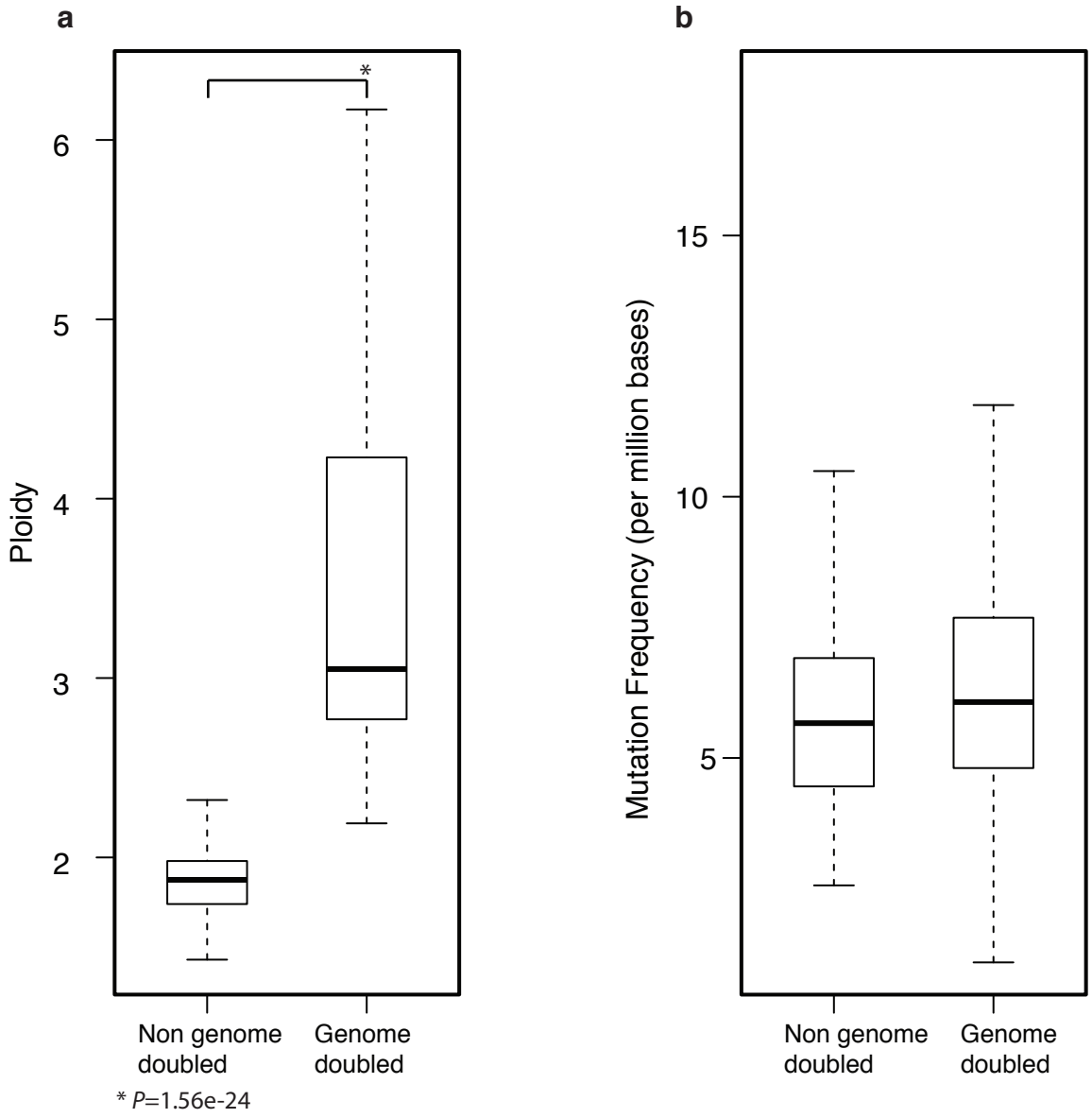
Average FISH centromeric probe count

CEP 2	CEP 4
2.2	2.14

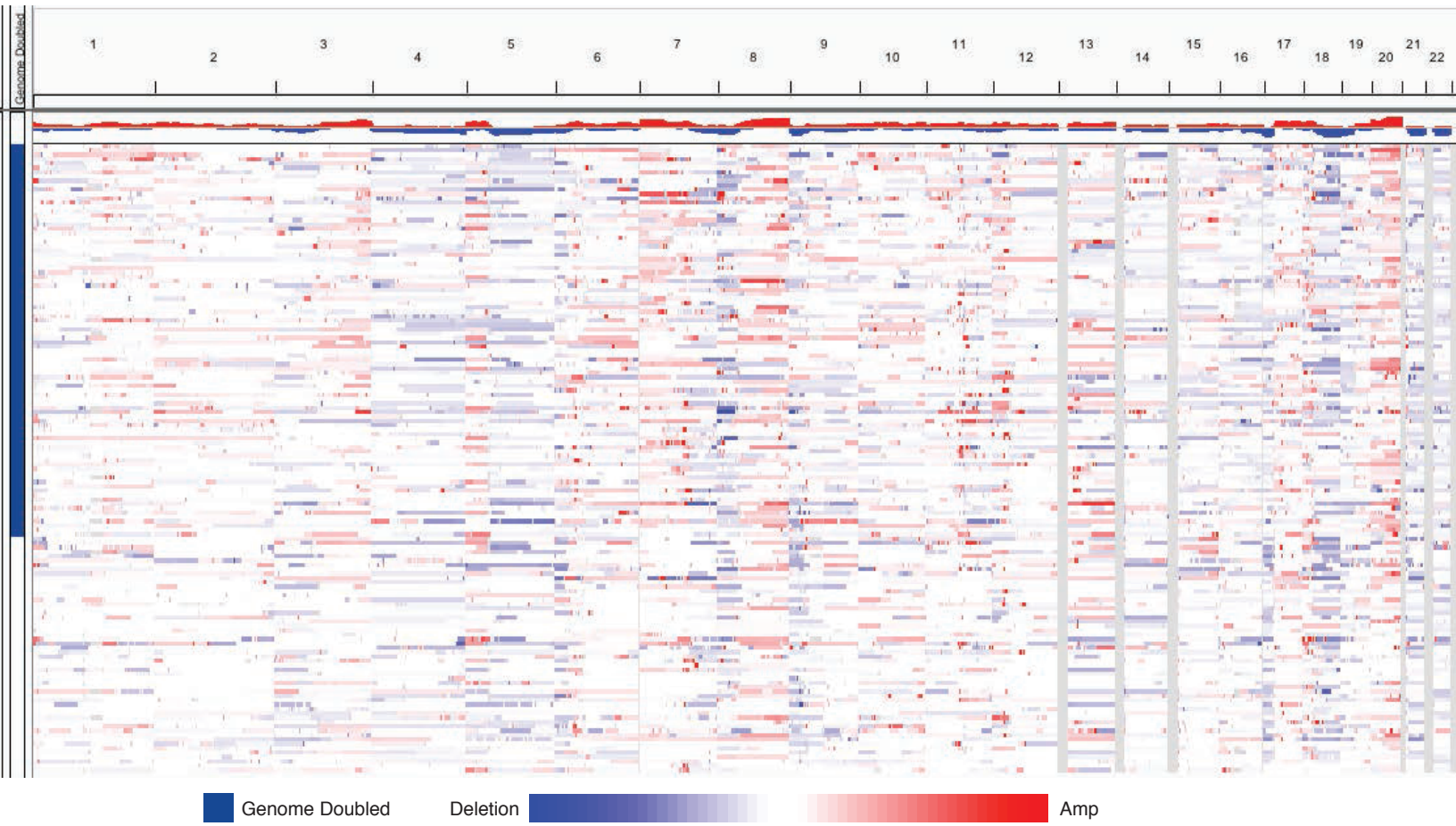
P = 0.01253

P = 0.0003110

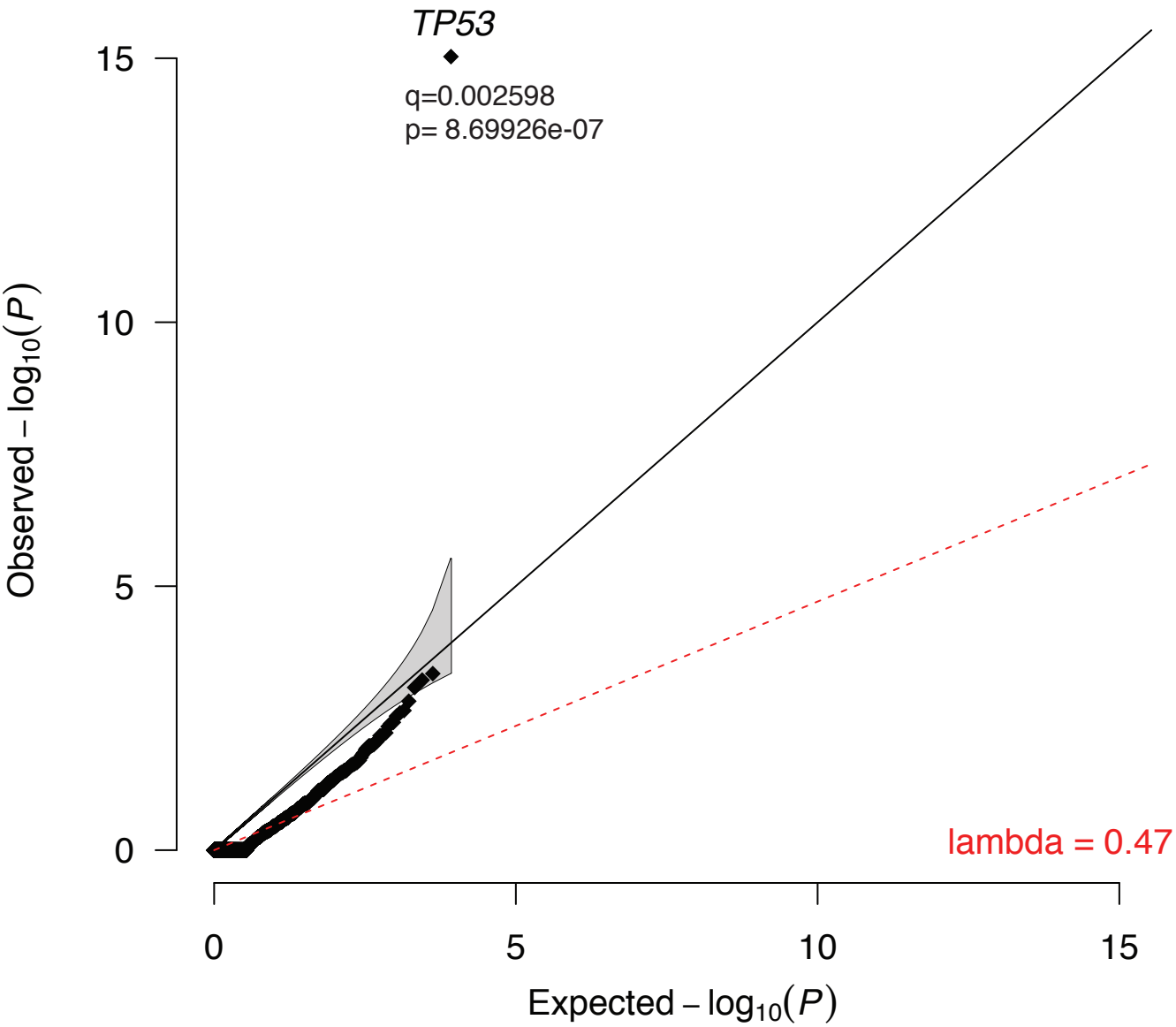
Supplementary Figure 16: Hematoxylin and Eosin stained sections of genomically and histologically distinct cancers isolated in patient P3.
Representative images of two foci of esophageal adenocarcinoma arising independently from each other. **a.**) EAC1 was non-genome doubled, contained a *CCND1* amplification, and had a mucinous histology. **b.**) EAC2 was genome doubled, contained amplifications in *MET*, *MYC*, and 3q26, and had a glandular and solid histology. Table below shows the number of enumerated probes for centromere 2 and 4.



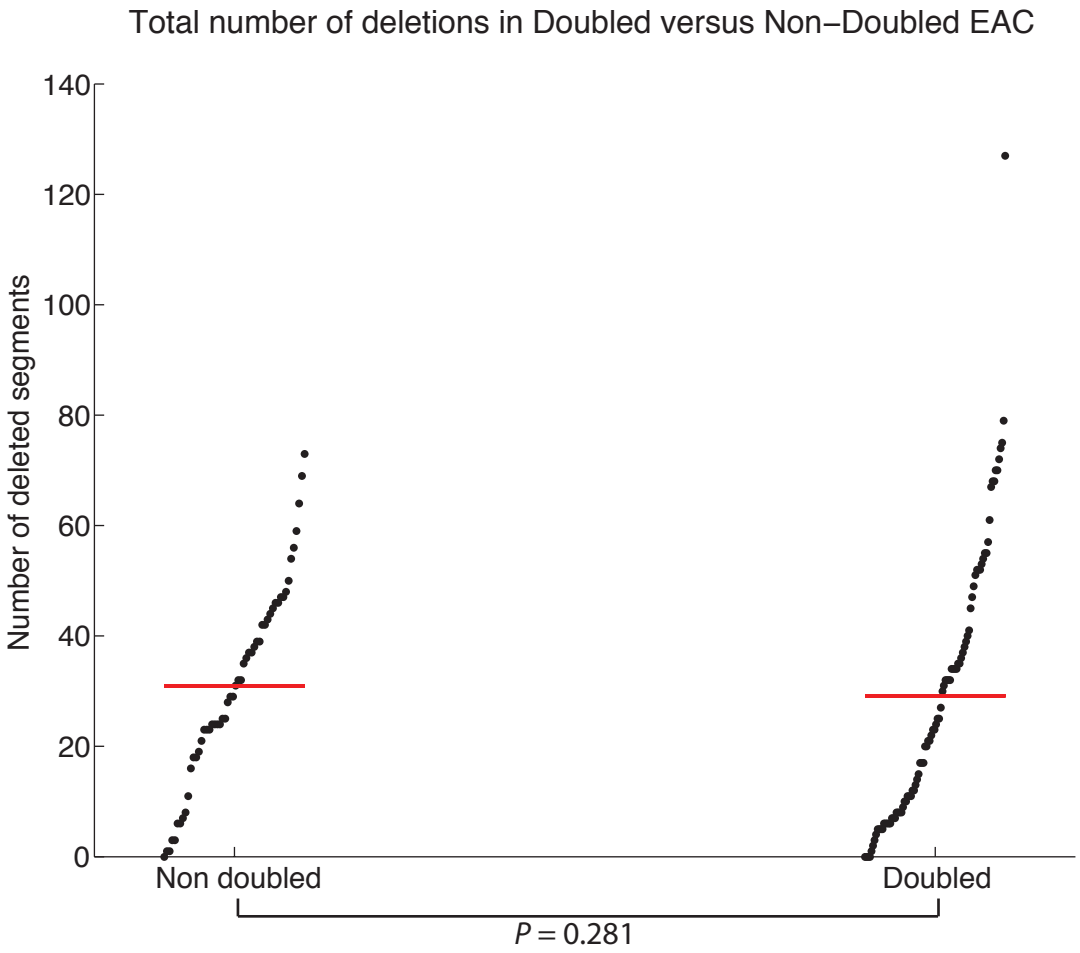
Supplementary Figure 17: Genome doubled EAC has a higher ploidy but equal mutation frequency compared to EAC without genome doubling.
The ABSOLUTE inferred **a**) ploidy and **b**) mutation rate in genome doubled (N=90) and non-genome doubled (N=54) EACs.

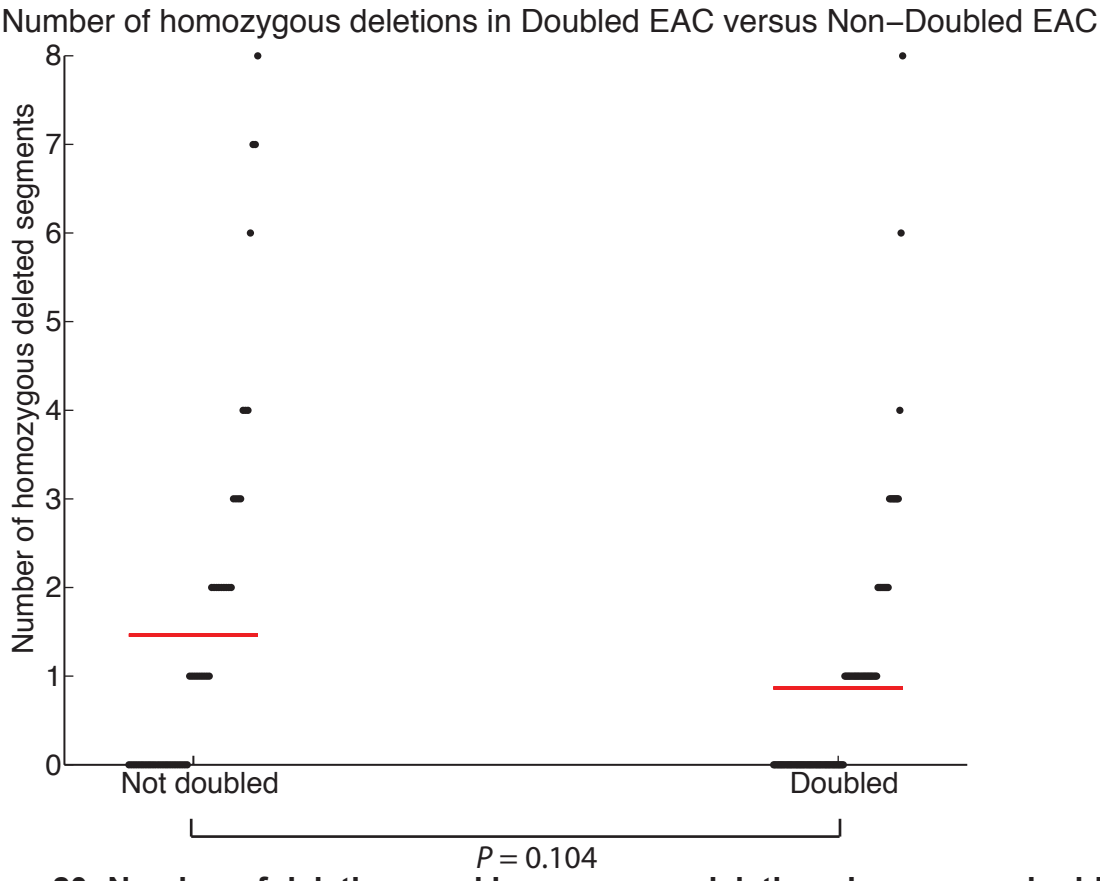


Supplementary Figure 18: EAC without genome doubling still contains frequent copy number alterations
Copy number heatmap of log2 ratios of the individual EAC samples separated by genome doubling status (genome doubled=colored bar on top right). Chromosome numbers listed on top right, blue=deletion, and red=amplification.



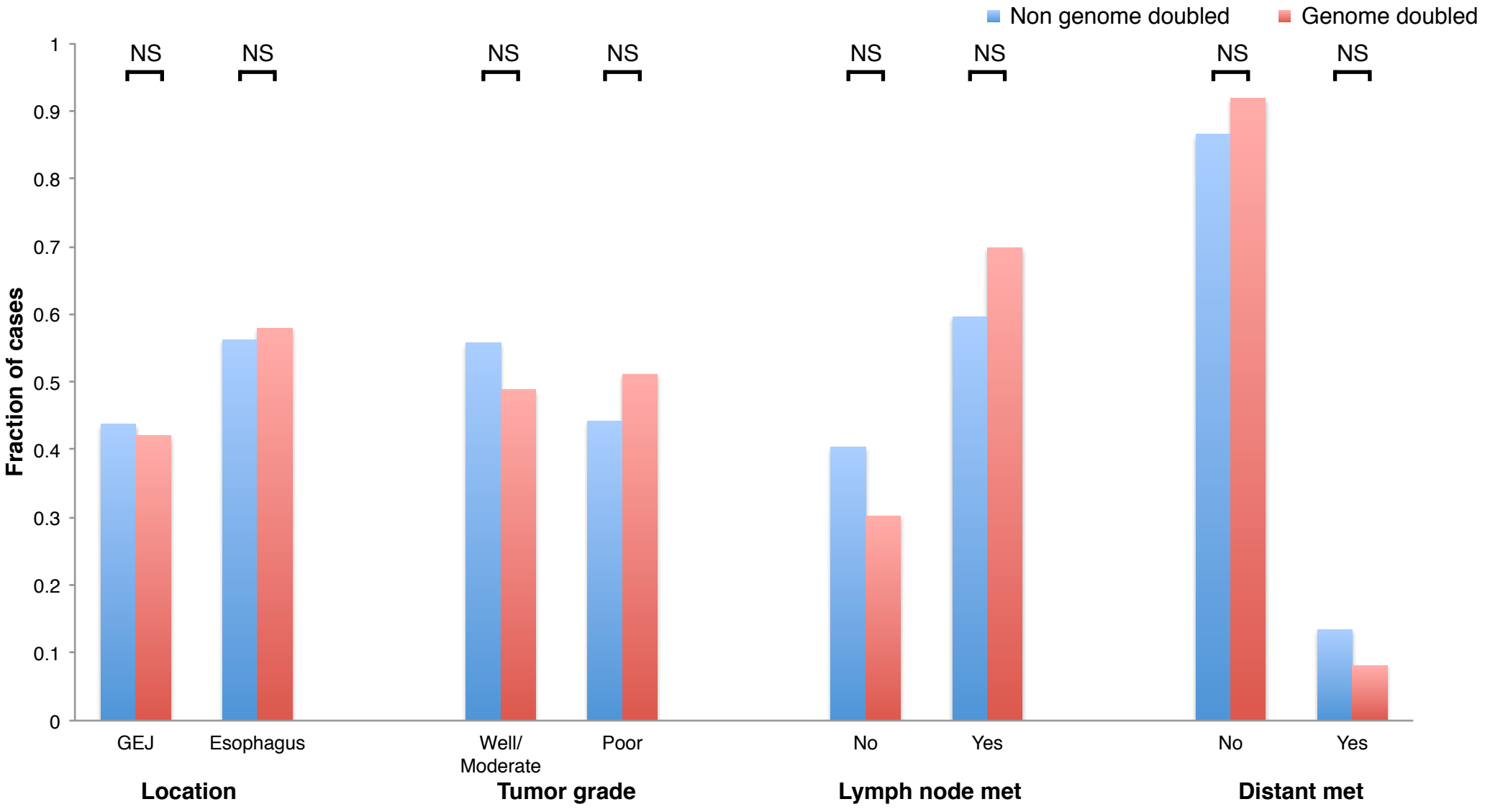
Supplementary Figure 19: TP53 mutation frequently occurs before genome doubling
Mutations that occur before a whole genome doubling (WGD) event will often have a multiplicity > 1 (as assessed by ABSOLUTE). We analyzed non-silent mutation multiplicities in the 90 genome-doubled EAC samples to search for genes whose mutations were significantly biased towards multiplicity values > 1. For each gene, we calculated the P-value against the null hypothesis in which having a mutation with multiplicity >1, in a given patient, was a random event that was independent of the gene. The statistic per gene was the number of patients in which it had a mutation with multiplicity > 1, and the P-values were calculated based on the tail of the distribution of the sum of Bernoulli variables (one per patient with mutation in the gene) with success probabilities pi representing the fraction of mutations with a multiplicity > 1 in the i-th patient. This analysis used the same procedure used in17 to detect significantly homozygously mutated genes. *TP53* was the only gene (out of 2,986 genes with > 0 non-silent mutations tested) whose mutations were statistically significantly more likely to have multiplicity > 1. The figure shows a QQ-plot, with individual gene –log10(P-values) represented with black diamonds; expected (uniform quantiles; x-axis) vs. observed (y-axis). The solid black line represents y=x. The shaded area indicates the 95% confidence interval of the uniform quantiles. The dashed red line represents a linear fit of y to x-1; the slope of 0.47 indicates that the statistical test is underpowered for many genes, as expected since most genes have only a small number of mutations. Q-values were calculated by correcting the observed P-values for multiple-hypothesis testing using Bonferroni correction.





Supplementary Figure 20: Number of deletions and homozygous deletions in genome doubled and non-genome doubled esophageal adenocarcinoma

Number of deleted (**a**) or homozygously deleted (**b**) segments per sample in genome doubled and non-genome doubled esophageal adenocarcinoma samples. Line represents the mean.



Supplementary Figure 21: Clinical characteristics in genome doubled and non-genome doubled esophageal adenocarcinoma.

Supplementary Figure 21: Clinical characteristics in genome doubled and non-genome doubled esophageal adenocarcinoma

Bar graphs showing the fraction of cases with the given clinical characteristics. No differences were observed between the to types of cancer.

Supplementary Table 3: Patient and tumor demographics for laser capture microdissected samples

Patient #	Age	Sex	Stage	Tumor Location	BE Length
P1	50	M	pT2N0	Esophagus	Long segment
P3	71	M	pT3N1b	Esophagus	1.5-3.7 cm
P4	75	F	pT1N0	Esophagus	4.5 cm
P6	64	M	pT3N1a	GEJ	short segment
P7	64	F	pT3N2	GEJ	5.0 cm

Supplementary Table 4: Tumor suppressor multiplicity

	# of mutations with Multiplicity	
Gene	**1**	**>1**
TP53	6	54
CDKN2A	4	3
SMAD4	0	1
ARID1A	2	5
ARID2	2	0
SMARCA4	3	1

Supplementary Table 5: Tumor suppressor gene and pathway loss

Pathway	Gene	No GD (54) # LOF* mut	frac	GD (90) # LOF* mut	frac	No GD (54) # Del	frac	GD (90) # Del	frac	No GD (54) #	total frac	GD (90) #	frac
	TP53	37	0.685	60	0.667	1	0.019	0	0.000	38	0.704	60	0.667
Chromatin Modifiers													
	ARID1A	3	0.056	7	0.078	2	0.037	0	0.000	5	0.093	7	0.078
	ARID1B	1	0.019	0	0.000	0	0.000	0	0.000	1	0.019	0	0.000
	ARID2	3	0.056	1	0.011	0	0.000	0	0.000	3	0.056	1	0.011
	PBRM1	2	0.037	0	0.000	3	0.056	0	0.000	5	0.093	0	0.000
	SMARCA4	3	0.056	4	0.044	1	0.019	1	0.011	4	0.074	5	0.056
	SMARCB1	1	0.019	1	0.011	1	0.019	0	0.000	2	0.037	1	0.011
Cell Cycle													
	CDKN1B	2	0.037	0	0.000	0	0.000	0	0.000	2	0.037	0	0.000
	CDKN2A	10	0.185	7	0.078	10	0.185	12	0.133	20	0.370	19	0.211
	CDKN2C	0	0.000	1	0.011	0	0.000	1	0.011	0	0.000	2	0.022
	RB1	0	0.000	1	0.011	1	0.019	0	0.000	1	0.019	1	0.011
TGFB													
	SMAD2	1	0.019	0	0.000	0	0.000	0	0.000	1	0.019	0	0.000
	SMAD4	11	0.204	1	0.011	5	0.093	2	0.022	16	0.296	3	0.033
WNT/B Catenin													
	SOX9	1	0.019	1	0.011	0	0.000	0	0.000	1	0.019	1	0.011
	APC	1	0.019	2	0.022	0	0.000	1	0.011	1	0.019	3	0.033
	AXIN1	2	0.037	0	0.000	0	0.000	1	0.011	2	0.037	1	0.011

*includes missense with cosmic >=3